AL-TR-1991-0153

AD-A251 053

ARMSTRONG LABORATORY

REVIEW OF PSYCHOPHYSICALLY-BASED IMAGE QUALITY METRICS

Jennie J. Gallimore, Ph.D.

WRIGHT STATE UNIVERSITY
DAYTON, OHIO

SEPTEMBER 1991

FINAL REPORT FOR PERIOD JUNE 1990 TO MAY 1991

Approved for public release; distribution is unlimited.

AIR FORCE SYSTEMS COMMAND
WRIGHT-PATTERSON AIR FORCE BASE, OHIO 45433-6573

NOTICES

TECHNICAL REVIEW AND APPROVAL

AL-TR-1991-0153

This report has been reviewed by the Office of Public Affairs (PA) and is releasable to the National Technical Information Service (NTIS). At NTIS, it will be available to the general public, including foreign nations.

This technical report has been reviewed and is approved for publication.

FOR THE COMMANDER

Kenneth R Boff

KENNETH R. BOFF, Chief
Human Engineering Division
Armstrong Laboratory

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	SEP 1991	FINAL June 1990 - May 1991

4. TITLE AND SUBTITLE

Review of Psychophysically-Based Image Quality Metrics

5. FUNDING NUMBERS

C: F33615-89-C-0532
PE: 62202F
PR: 7184
TA: 10
WU: 44

6. AUTHOR(S)

Jennie J. Gallimore, Ph.D.
Wright State University

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Logicon Telchnical Services, Inc.
P.O. Box 317258
Dayton, OH 45431-7258

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Armstrong Laboratory
Crew Systems Directorate
Human Engineering Division
HSD AFSC
Wright-Patterson Air Force Base, OH 45433-6573

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

AL-TR-1991-0153

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

A

13. ABSTRACT *(Maximum 200 words)*

There have been many advances in imaging technologies resulting in a variety of techniques for generation, coding for transmission, image processing, decoding, and display of information. In addition to the development of techniques, investigators have recognized the need for quantitative measures of image distortion and/or image quality corresponding to observer performance and observer impressions of the images. The purpose of this report is to review current psychophysically based measures of image quality for possible application to compressed or transmitted sensory imagery. This report briefly reviews human vision models, image quality metrics, empirical modelling approaches to image quality, discusses current research needs and provides recommendations for immediate application to compressed digital images.

14. SUBJECT TERMS

Image Quality, Display, Modulation Transfer Function

15. NUMBER OF PAGES

39

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UNLIMITED

GENERAL INSTRUCTIONS FOR COMPLETING SF 298

The Report Documentation Page (RDP) is used in announcing and cataloging reports. It is important that this information be consistent with the rest of the report, particularly the cover and title page. Instructions for filling in each block of the form follow. It is important to ***stay within the lines*** to meet ***optical scanning requirements***.

Block 1. Agency Use Only *(Leave blank)*.

Block 2. Report Date. Full publication date including day, month, and year, if available (e.g. 1 Jan 88). Must cite at least the year.

Block 3. Type of Report and Dates Covered. State whether report is interim, final, etc. If applicable, enter inclusive report dates (e.g. 10 Jun 87 - 30 Jun 88).

Block 4. Title and Subtitle. A title is taken from the part of the report that provides the most meaningful and complete information. When a report is prepared in more than one volume, repeat the primary title, add volume number, and include subtitle for the specific volume. On classified documents enter the title classification in parentheses.

Block 5. Funding Numbers. To include contract and grant numbers; may include program element number(s), project number(s), task number(s), and work unit number(s). Use the following labels:

C	- Contract	**PR**	- Project
G	- Grant	**TA**	- Task
PE	- Program Element	**WU**	- Work Unit Accession No.

Block 6. Author(s). Name(s) of person(s) responsible for writing the report, performing the research, or credited with the content of the report. If editor or compiler, this should follow the name(s).

Block 7. Performing Organization Name(s) and Address(es). Self-explanatory.

Block 8. Performing Organization Report Number. Enter the unique alphanumeric report number(s) assigned by the organization performing the report.

Block 9. Sponsoring/Monitoring Agency Name(s) and Address(es). Self-explanatory.

Block 10. Sponsoring/Monitoring Agency Report Number. ***(If known)***

Block 11. Supplementary Notes. Enter information not included elsewhere such as: Prepared in cooperation with...; Trans. of...; To be published in.... When a report is revised, include a statement whether the new report supersedes or supplements the older report.

Block 12a. Distribution/Availability Statement. Denotes public availability or limitations. Cite any availability to the public. Enter additional limitations or special markings in all capitals (e.g. NOFORN, REL, ITAR).

DOD - See DoDD 5230.24, "Distribution Statements on Technical Documents"
DOE - See authorities.
NASA - See Handbook NHB 2200.2.
NTIS - Leave blank.

Block 12b. Distribution Code.

DOD - Leave blank.
DOE - Enter DOE distribution categories from the Standard Distribution for Unclassified Scientific and Technical Reports.
NASA - Leave blank.
NTIS - Leave blank.

Block 13. Abstract. Include a brief *(Maximum 200 words)* factual summary of the most significant information contained in the report.

Block 14. Subject Terms. Keywords or phrases identifying major subjects in the report.

Block 15. Number of Pages. Enter the total number of pages.

Block 16. Price Code. Enter appropriate price code *(NTIS only)*.

Blocks 17. - 19. Security Classifications. Self-explanatory. Enter U.S. Security Classification in accordance with U.S. Security Regulations (i.e., UNCLASSIFIED). If form contains classified information, stamp classification on the top and bottom of the page.

Block 20. Limitation of Abstract. This block must be completed to assign a limitation to the abstract. Enter either UL (unlimited) or SAR (same as report). An entry in this block is necessary if the abstract is to be limited. If blank, the abstract is assumed to be unlimited.

PREFACE

This effort was performed in support of Air Force Work Unit 7184 10 44, "Advanced Strategic Cockpit Engineering and Research." Mr Gilbert Kuperman, Crew Station Integration Branch, Human Engineering Division, Directorate of Crew Systems, Armstrong Laboratory, Wright-Patterson Air Force Base, Ohio, was the Work Unit Manager and provided technical guidance for the effort. The effort was carried out as a subcontract to Dr Gallimore from Logicon Technical Services, Inc., Dayton, Ohio, under Air Force Contract Number F33615-89-C-0532. Mr Robert Linhart was the contract monitor.

Table of Contents

SECTION PAGE

I. INTRODUCTION 1

II. HUMAN VISION MODELS 3
Contrast Threshold Function 3
Weber-Fechner Fraction and Steven's Power Law 8
Model of Monochromatic Vision 9
Multichannel Spatial Frequency Models 10

III. IMAGE QUALITY METRICS 13
MTF Based Metrics 13
Modulation Transfer Function Area 14
Gray Shade Frequency Product 17
Integrated Contrast Sensitivity Function 18
Visual Capacity 20
Information Content 21
Signal-to-Noise 23
Application to Digitally Compressed Imagery 24
Pixel-Based Metrics 26
Mean Square Error 27
Perceptual Mean Square Error 27
Contrast Energy Difference Metric 28

IV. EMPIRICAL MODELLING 31

V. HUMAN PERFORMANCE MEASURES 33

VI. RECOMMENDATIONS 35

List of Tables

Table 1. Correlations between MTFA and task performance measures 16

Table 2. Correlations between GSFP and task performance measures 19

Table 3. Correlations between ICS and task performance measures 19

Table 4. Correlations between VC and EP metrics with task performance measures 21

Table 5. Correlations between IC and task performance measures 23

Table 6. Correlations between SN and task performance measures 24

Table 7. Correlations between image dependent metrics evaluated by Beaton (1984) and subjective rating performance 26

List of Figures

Figure 1. The modulation transfer function 5

Figure 2. The contrast threshold function 6

Figure 3. Monochromatic visual model 11

Figure 4. The modulation transfer function area concept 15

SECTION I - REVIEW OF PSYCHOPHYSICALLY BASED IMAGE QUALITY METRICS

INTRODUCTION:

There have been many advances in imaging technologies resulting in a variety of techniques for generation, coding for transmission, image processing, decoding, and display of information. The purpose of these systems is to generate images such that observers can extract relevant information; therefore, it is necessary to consider human observer requirements. Models of the human visual system have been used during development of the variety of techniques referred to above. However, in addition to the development of techniques, investigators have recognized the need for quantitative measures of image distortion and/or image quality corresponding to observer performance and observer impressions of the images. Quantitative measures of image quality have the potential for reducing the need for multiple experiments to test the variety of image processing techniques that may be applied to an image, but still enable verification of the technique in terms of human requirements. The purpose of this report is to review current psychophysically based measures of image quality for possible application to compressed or transmitted sensor imagery.

Although some image quality metrics have been based only on physical measures of the image, many are based on models of the human visual system. Section II of this report briefly introduces human visual models that are often used in image quality metrics. Section III describes many image quality metrics including research results of studies using these metrics. An additional important consideration for use of metrics is the type of performance measures that are to be correlated with the metric. The metrics are developed in hopes of being highly correlated with human performance and/or perception; however, research has not focused on the need for investigating task performance measures. Section IV discusses this issue. Recommendations for application of metrics to digitally compressed imagery are summarized in Section V.

SECTION II - HUMAN VISION MODELS

Many image quality metrics have been based only on physical measures of the image and do not take into consideration the workings of the human visual system. It is therefore not surprising that these metrics do not correlate well with human performance. Many vision models deal with representation of the human response to spatial inputs, such as static spatial variation in luminance. Spatial frequency (SF) models are the primary models used in current image quality metrics. It should be noted that human vision models have been developed for early stages of vision and do not take into consideration higher levels of cognitive processing. This section discusses approaches to modelling the human visual system. Application of these models to image quality metrics will follow in the succeeding section.

Contrast Threshold Function (CTF):

Linear systems analysis and the mathematics of Fourier transforms have been applied to the analysis of imaging systems to determine the modulation transfer function (MTF) of the system. Modulation is defined as:

$$M = \frac{L_{max} - L_{min}}{L_{max} + L_{min}} \quad (1)$$

where L_{max} is the maximum luminance and L_{min} is the minimum luminance of a sinusoidal signal. With linear systems analysis, it is possible to determine the extent to which any component or system of components can transmit a signal. During transmission, some of the signal is lost due to limitations in the system. The modulation transfer factor is the ratio of the modulation out of the system to the modulation into the system,

$$T(\omega,\nu)=\frac{M_o(\omega,\nu)}{M_I(\omega,\nu)} \tag{2}$$

where T (ω, v) is the modulation transfer factor at spatial frequencies ω, v and M_o and M_I are the output and input modulations respectively. If the modulation transfer factor values at each spatial frequency are connected, a continuous function is formed termed the modulation transfer function (MTF). The loss of output modulation generally increases with increasing frequency of the sine wave input (see Figure 1).

The concepts of linear systems analysis have been applied to the visual system. An observer is presented with a known sine wave pattern that is varied in spatial frequency and is asked to adjust the luminance modulation of the grating to visual threshold. When results are plotted as a function of spatial frequency, the function is termed the contrast threshold function (CTF) as illustrated in Figure 2. This technique can be considered a "black box" approach to modelling the visual response.

To fit experimental data, Dooley (1975, cited by Levine, 1985) developed the following equation for the CTF,

$$CTF(\omega)=5.05(e^{-0.138\omega})(e^{0.1\omega}) \tag{3}$$

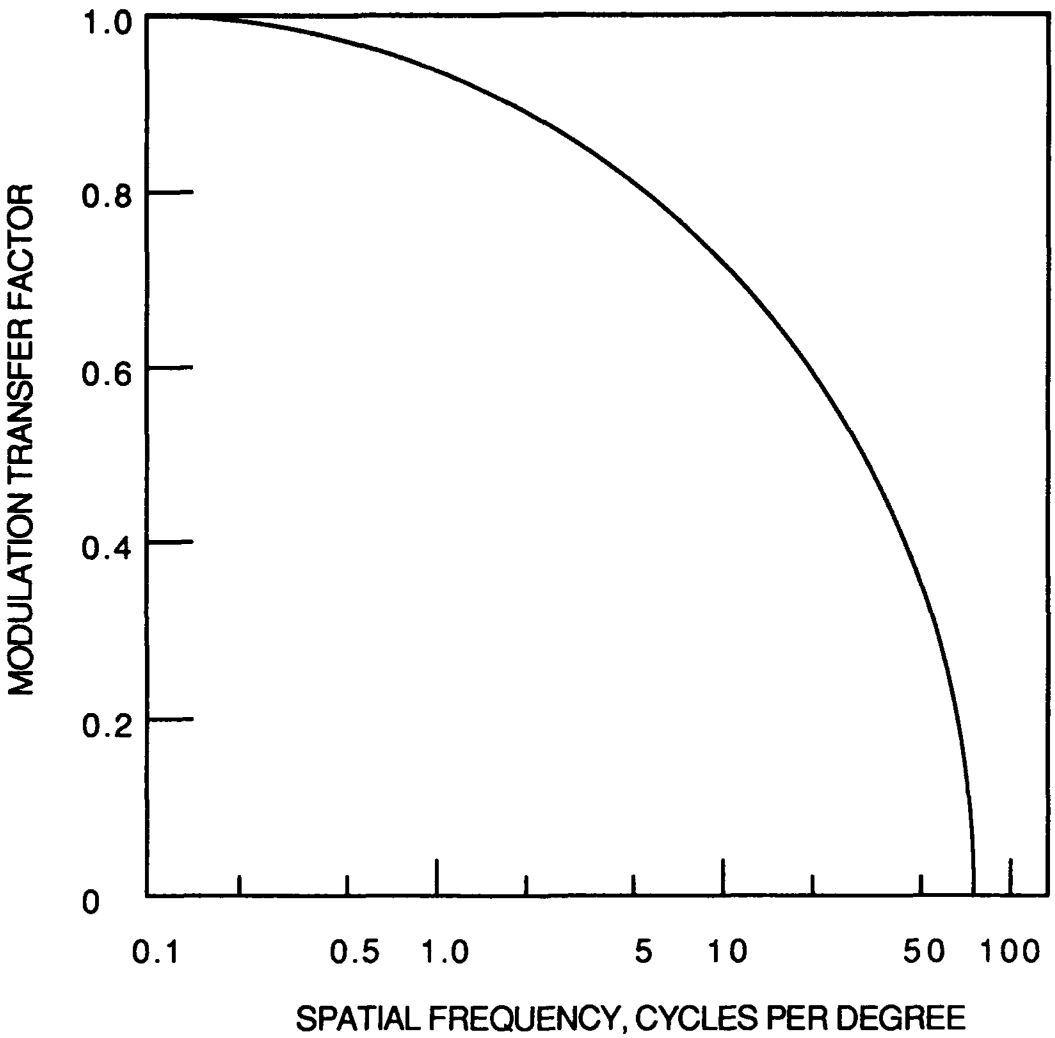

Figure 1: The modulation transfer function (MTF).

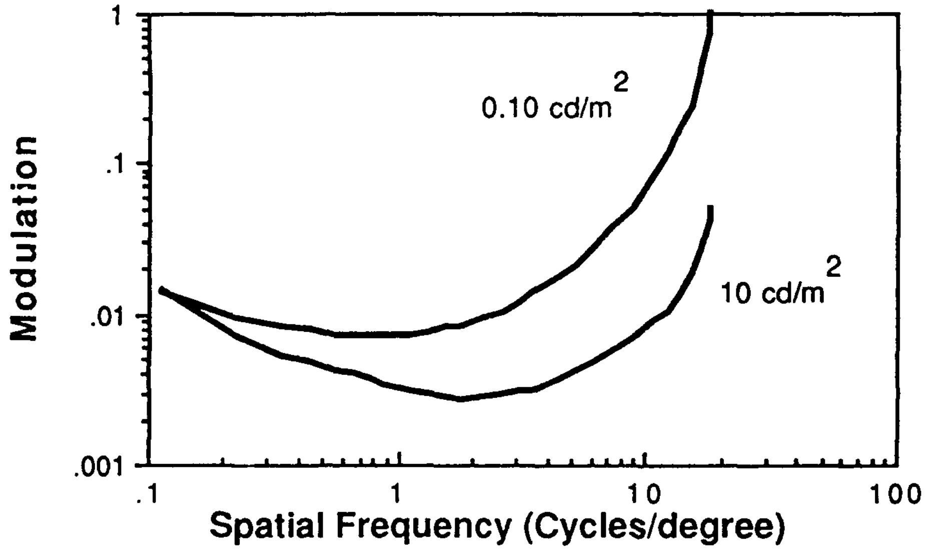

Figure 2: The contrast threshold function (CTF).

where ω is the spatial frequency in cycles per degree of visual angle. This equation has been normalized to the peak modulation of 0.005.

Beaton (1988) also provides an equation for the CTF based on experimental data for young eyes and viewing distances greater than 18 inches,

$$CTF(\omega) = b_0 e^{b_1\omega + b_2\omega^2 + b_3\omega^3} \tag{4}$$

where,

$b_0 = 1.7062 \times 10^{-3}$

$b_1 = 201.6188 \times 10^{-3}$

$b_2 = -2.31616 \times 10^{-3}$

$b_3 = 0.20000 \times 10^{-6}$

Research results have illustrated that the CTF will shift as visual parameters are changed. Variables which cause shifts in the CTF include display luminance, orientation of the grating, type of wave pattern, and viewing distance, to name a few. (For a review of this literature readers are referred to Snyder, 1980). Because the CTF varies as a function of different parameters, it is important to use a CTF that is similar to the viewing situation. Therefore, researchers may be required to empirically determine the CTF for the specific viewing situations.

Results of experiments with the CTF illustrate that the human visual system does not meet linear systems analysis assumptions of linearity and isotropy. However, linear systems analysis will work if the response is considered at least linear in the range investigated. It has also been argued that the CTF is a threshold measure and may not apply well to suprathreshold levels that humans deal with in real-life situations. However, Ginsburg, Cannon, and Nelson (1980) and Decker (1989) illustrated that CTF can represent suprathreshold processing as well. Given the above assumptions, the CTF approach is feasible and has resulted in good correlations with performance when used with image quality metrics.

Weber-Fechner Fraction and Steven's Power Law:

The visual response to intensity is nonlinear, and the nonlinearity is thought to take place at the photoreceptor level of processing. Psychophysical models of the nonlinear response have been investigated for many years. In 1886, Weber presented subjects with a background of intensity I and a target against the background of intensity I + ΔI. Subjects were instructed to determine when they could just detect a difference between the background and target (just noticeable difference, JND). Weber found that the proportion by which the stimulus I must increase in order to just detect the difference was a constant such that,

$$K=\frac{\Delta I}{I} \tag{5}$$

This formula is termed the Weber fraction (Coren, Porac, and Ward, 1984). However, this linear equation does not hold for low or high intensity values. Experiments by Fechner led to a change in the Weber fraction that indicated a logarithmic response and is termed the Weber-Fechner fraction.

$$K=\frac{\Delta I}{\log I} \tag{6}$$

This fraction indicates that it requires a small physical change to achieve one JND for a weak stimulus and a larger change to achieve one JND for a stronger stimulus. The logarithmic relationship has long been included in many models of human vision.

Stevens (1961, cited by Coren et. al., 1984) proposed a power law to explain psychophysical responses to stimuli. The human response is related to input intensity (I) as,

$$S=K(I-I_o)^n \tag{7}$$

where S is the sensation or response, K is a constant, I_o is absolute intensity at threshold, and n is an exponent which varies depending on the sensory input (e.g., hearing, visual, tactile). If $n < 1$, the curve is concave downward and indicates that the more intense a stimulus the greater that stimulus must be changed to produce the same response. The psychophysical response to brightness results in exponents less than 1.

Model of Monochromatic Vision:

Hall and Hall (1977) developed a model of the visual system to match the results of the contrast sensitivity tests for monochromatic vision. Their model is based on models originally discussed by Stockham (1972) and Mannos and Sakrison (1974).

The model is composed of three subsystems as illustrated in Figure 3. The first subsystem represents the ocular optical system and is a low pass filter defined as

$$H_1(\omega)=\frac{2\alpha}{\alpha^2+\omega^2} \tag{8}$$

where, $\alpha=\pi\Delta$ is the spatial angular frequency. The value of alpha depends upon the pupil diameter. For a white light and pupil diameter of 3 mm, α = 0.7.

The second subsystem describes the nonlinearity of the visual system. Hall and Hall used a logarithmic process. Mannos and Sakrison (1974) proposed a power function at this stage of the model.

The third subsystem is the high pass filter. It is employed to take into consideration lateral inhibition and is defined by the following equation:

$$H_2(\omega)=\frac{a^2+\omega^2}{2a_o a+(1-a_o)(a^2+\omega^2)} \tag{9}$$

where a_o is a constant distance factor relating to the distance between photoreceptors and a is a strength of inhibition factor. Hall and Hall set parameters a_o = 0.01 and a = 0.2 to match the CTF results reported by Davidson (1968, as cited by Hall and Hall, 1977).

Multichannel Spatial Frequency Models:

Multichannel models assume that the visual system is composed of multiple

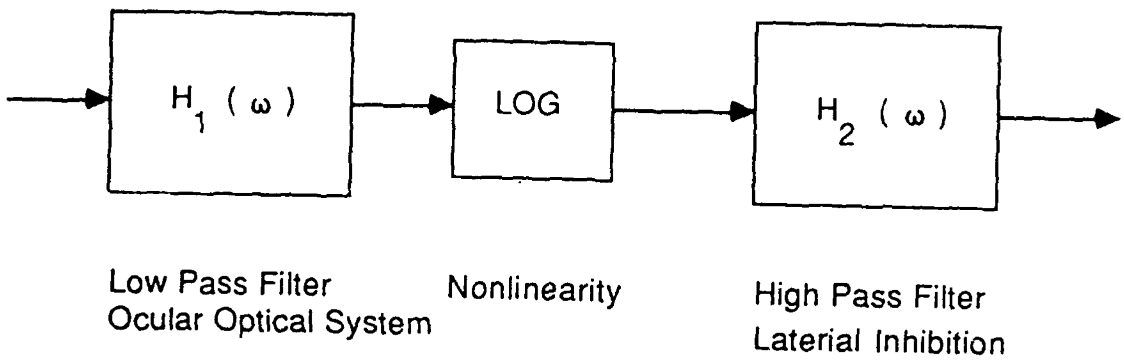

Figure 3: Monochromatic visual model (from Hall and Hall, 1977).

independent narrow-band channels. Each channel consists of a collection of cells over the retinal field. Each channel has a specific bandwidth and center frequency such that each channel is sensitive to a different range of spatial frequencies. The bandwidths are expressed in octaves. When the channels are pooled, results predict the CTF of human vision. (Wilson and Bergen, 1979). Currently, the multi-channel vision model has the most support and a great deal of research has been published specifying these channels.

SECTION III - IMAGE QUALITY METRICS

There are different criteria for evaluating an image or for evaluating the techniques for generating or processing an image. Criteria may include length of processing time, or computational resources required. Some metrics are developed to measure the physical differences between an original image and a processed image. These types of criteria or metrics do not take into consideration how the observer will perform when viewing the image and are sometimes referred to as image fidelity measures. This section will discuss metrics that have been developed based on psychophysical techniques which incorporate the visual models described previously.

MTF Based Metrics:

This section describes metrics which were originally developed for continuous tone film images and were later applied to cathode ray tube (CRT) images. These metrics have been described fully by Task (1979), Beaton (1984) and Decker, Pigion, and Snyder (1989). The metrics described are those which take into consideration the visual system or have been behaviorally validated.

Many studies investigating metrics concentrate on one metric and report results. Comparison of metrics across studies is not always possible because of experimental differences. However Task (1979) and Beaton (1984) conducted research comparing a variety of metrics using the same imagery for each metric.

Task (1979) compared metrics for film and video images using three types of target detection and recognition studies. In this research, the quality of the image was changed by changing the system MTF. Beaton (1984) also examined a variety of metrics for hard copy images as well as CRT displayed images. In this research, digital images were used and were degraded by blur and noise. Two tasks were employed for photointerpreters, a subjective rating scale task, and an information extraction task. Results from studies conducted by Task and Beaton will be reported as each metric is discussed. Additional metrics from other sources will also be discussed.

Modulation Transfer Function Area (MTFA):

The modulation transfer function area (MTFA) metric combines the MTF of an imaging system and the visual contrast threshold function (CTF). This metric has received much attention and research results indicate that this metric correlates well with performance. The MTFA can be described as the area between the zero spatial frequency and the crossover frequency of the two curves. The crossover frequency is the "limiting resolution." The MTFA is illustrated in Figure 4. The MTFA is defined as,

$$MTFA=\Delta\omega\Delta\nu\sum_{\omega=-f}^{\omega=f}\sum_{\nu=-f}^{\nu=f}[T_s(\omega,\nu)-T_e(\omega,\nu)] \tag{10}$$

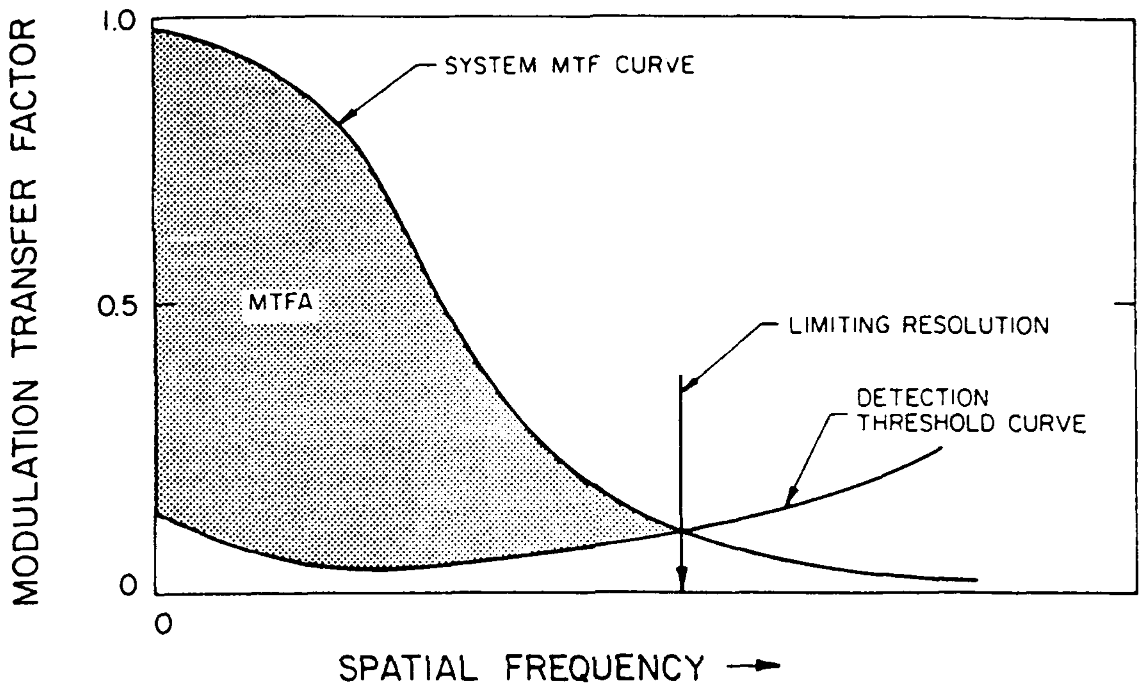

Figure 4: The modulation transfer function area (MTFA) concept (From Decker, Pigion, and Snyder, 1989).

where, $T_s(\omega, v)$ is the composite MTF of the imaging system. $T_e(\omega, v)$ is the CTF, and f is the spatial frequency where the MTF and CTF have the same value or the "limiting resolution."

Both Beaton (1984) and Task (1979) evaluated the MTFA. Table 1 summarizes the correlations between the MTFA metric and the different performance measures for both studies.

Table 1. Correlations between MTFA and task performance measures.

Display Media	Performance Measure	MTFA Correlation (R^2)
Film	angle subtended at recognition	-0.95 (log MTFA)
CRT	angle subtended at recognition	-0.878 (log MTFA)
CRT	slant range at detection	0.866 (log MTFA)
CRT*	Information extraction	0.79

where CRT* is digitally addressed CRT.

Note that the MTFA metric gives the highest correlation for film. Other studies have shown correlations ranging from 0.211 to 0.97 for CRT displayed images (Decker, et al., 1987).

More recently Task and Pinkus (1987) evaluated the MTFA using a target detection task. They varied the MTF of the video display system to include low and high contrast conditions. The CTF's for each subject were measured at 8 spatial frequencies. The stimuli were presented using 16mm motion picture film displayed on a white phosphor monochrome television display. Correlations between MTFA and detection recognition for low and high contrast conditions were -0.269 and 0.005 respectively. Correlations for combined low and high contrast conditions were -0.575. This study indicated a lack of correlation. However, it should be noted that stimuli were not static images. The target was zoomed at a fixed rate until the subject recognized the target. A metric for dynamic image quality may be needed in this case.

Gray Shade Frequency Product (GSFP):

This metric was proposed by Task and Verona (1976). The MTFA assumes that the excess MTF over the CTF is isotropic for all spatial frequencies and all modulations above the threshold CTF. Beamon and Snyder (1975) suggested that the area just above the CTF is more important to the observer because it is important to have a modulation above the minimal required but increases in excess MTF are not important in most tasks. The GSFP is a nonlinear transform of the MTFA to weight the area near the CTF more heavily. This metric models the visual system as a logarithmic amplifier such that the visual system "sees" modulations proportional to the logarithm of the modulation. Modulation is transformed into "shades of gray" G, as follows:

$$G = 1 + \frac{\log_{10}[(1+M)/(1-M)]}{\log_{10}(2.0^{0.5})} \qquad (11)$$

where the numerator is the modulation and the denominator is the modulation between successive shades of gray. It should be noted that the denominator does not represent a psychophysical "just noticeable difference" between luminance levels which would perhaps be more appropriate.

The GSFP is defined as:

$$GSFP = \Delta\omega\Delta\nu \sum_{\omega=-f}^{\omega=f} \sum_{\omega=-f}^{\omega=f} G[T_s(\omega,\nu) - T_e(\omega,\nu)] \qquad (12)$$

Table 2 summarizes the correlations between the GSFP metric and different performance measures. GSFP does not appear to be advantageous over the MTFA.

Integrated Contrast Sensitivity Function (ICS):

The integrated contrast sensitivity function (ICS) was proposed by van Meeteren (1973). This metric simply weights the MTF of the system by the contrast sensitivity function (CSF, the inverse of the CTF) for each spatial frequency.

$$ICS = \Delta\omega\Delta\nu \sum_{\omega=-f}^{\nu=f} \sum_{\omega=-f}^{\omega=f} T_s(\omega,\nu) C_i(\omega,\nu) \qquad (13)$$

where $C_i(\omega, v)$ is the inverse of the CTF $(T_e(\omega, v))$.

Results of correlations between ICS and performance are summarized in Table 3.

Table 2. Correlations between GSFP and task performance measures.

Display Media	**Performance Measure**	**Correlation (R^2)**
Film	angle subtended at recognition	-0.858 (log GSFP)
CRT	angle subtended at recognition	-0.847 (log GSFP)
CRT	slant range at detection	0.869
CRT	information extraction	0.80
CRT	subjective ranking	0.73

Table 3. Correlations between ICS and task performance measures.

Display Media	**Performance Measure**	**Correlation (R^2)**
Film	angle subtended at recognition	-0.978
CRT	angle subtended at recognition	-0.818
CRT	information extraction	0.95
CRT	subjective ranking	0.95

The correlations for the ICS metric are higher than the MTFA correlations which is not unexpected. van Meeteren suggested this metric to be more sensitive to small changes in the MTF or CSF because multiplication is being used. That is, MTFA and GSFP subtract the CTF whereas this metric multiplies the values.

Visual Capacity (VC):

The visual capacity metric was introduced by Cohen and Gorog (1974). It is based on Schade's equivalent passband metric proposed in 1953. The original EP metric was defined as

$$EP=\Delta\omega\Delta\nu\sum_{x=o}^{n-1}\sum_{y=o}^{m-1}[T_s(\omega,\nu)]^2 \tag{14}$$

EP is the equivalent bandwidth of a rectangular MTF containing the same total sine-wave power as the actual MTF of the imaging system being measured. In other words, it is the cut-off frequency of a perfect filter passing the same power. The EP metric is related to the "sharpness" of an image or the width of the edge transitions in the image. VC is defined as

$$VC=A\Delta\omega\Delta\nu\sum_{x=0}^{n-1}\sum_{y=o}^{m-1}T_s(\omega,\nu)^2T_e(\omega,\nu)^2 \tag{15}$$

where A denotes the area of the display device and is used to normalize the metric to express the maximum number of perceived edge transitions. This metric is designed express the perceptual width of the edge transitions, taking into consideration the CTF (T_e). Beaton (1984) evaluated this metric and results are summarized in Table 4. The table also includes correlations of performance with the EP metric. Results indicate that including the CTF in the metric results in higher correlations.

Table 4. Correlations between VC and EP metrics with task performance measures.

Display Media	Performance Measures	Correlations (R^2)
	VC	
CRT	Information Extraction	0.87
CRT	Subjective Ranking	0.90
	EP	
CRT	angle subtended at recognition	-0.726
CRT	slant range at detection	0.761
CRT	Information Extraction	0.78
CRT	Subjective Ranking	0.69

Information Content (IC):

Schindler (1976) used the concept of information theory in development of this metric. IC is defined as,

$$IC = \Delta\omega\Delta\nu \sum_{x=0}^{n-1} \sum_{y=o}^{m-1} \log_2[1 + \frac{T_s(\omega,\nu)}{T_d(\omega,\nu)}] \tag{16}$$

where T_d refers to the "just-detectable" response level of the imaging system. Beaton replaced T_d with the CTF (T_e). Results are summarized in Table 5. It would be interesting to utilize this equation using just noticeable difference responses to digitally compressed images.

An interesting alternative to using the CTF (T_e) in this equation would be to use Just-noticeable difference responses or magnitude estimation responses to digitally compressed images. These are psychophysical techniques for determining difference thresholds, or "just detectable" responses from human observers. The JND or magnitude estimation function would be empirically derived.

For the JND technique, observers are presented with a range of digitally compressed images, for example 1 to 8 bits/pixel. One image is presented as the standard. As images are presented to the subject, they are asked to determine if the image is "better" or "worse" than the standard in terms of image quality. A function is plotted which indicates the proportion of "better" responses for each comparison image. This function could be used in place of (T_d).

The magnitude estimation approach is very similar. Subjects are asked to assign a number to an image based on a dimension of the stimuli. In this example, subjects could provide a subjective estimation of "quality" of the image, or noise in the image compared to a standard image. A function is plotted which indicates the magnitude estimations as a function of compression.

The difference with this approach as compared to using the CTF is that the CTF is based on detection of sinusoidal patterns. If a variety of images are used, the cognitive component is included in the subject's subjective impression of image content. However, the output functions with this approach are not expressed in spatial frequency; therefore, the equation would have to be modified.

Table 5. Correlations between IC and task performance measures.

Display Media	Performance Measure	Correlation (R^2)
CRT	Information Extraction	0.86
CRT	Subjective Ranking	0.84

Signal-to-Noise (SN):

Noise has been shown to effect visual performance. Beaton (1984) defined a signal-to-noise (SN) metric based on a metric by Hufnagel (1965). (See Beaton, 1984 for a discussion of Hufnagel's metric.) SN is defined as,

$$SN=\frac{\Delta\omega\Delta\nu\sum_{x=0}^{n-1}\sum_{y=0}^{m-1}T_s(\omega,\nu)^2T_e(\omega,\nu)^2}{[\Delta\omega\Delta\nu\sum_{x=0}^{n-1}\sum_{y=0}^{m-1}W_i(\omega,\nu)T_e(\omega,\nu)^2]^{0.5}} \tag{17}$$

where W_i is the Weiner noise power spectrum. The Weiner noise power spectrum is weighted by the CTF. The denominator represents the root mean square deviation of the perceptually weighted noise signal. This is the only metric which actually directly measures display noise rather than determining a separate CTF for each noise condition. Beaton compared this metric with 16 others and the SN metric yielded the highest correlations with performance. Table 6 summarizes Beaton's (1984) research findings using this metric.

Table 6. Correlations between SN and task performance measures.

Display Media	Performance Measure	Correlations
CRT	Information Extraction	0.95
CRT	Subjective Ranking	0.95

Application to Digitally Compressed Imagery:

The image quality metrics discussed above were developed to determine the image capability of the display (or system) and the metrics are image independent. Assuming that a display system is capable of exceeding the capabilities of human vision, then the need for metrics which quantify the display system would not be necessary. However, in addition to determining the quality of an imaging device, it is also important to determine the effect of image processing (such as compression) on image quality. The metrics described above do not account for the image processing technique. However, it is possible to change each metric to an image dependent metric by using the display modulation spectrum of the image,

$$M_o(\omega,\nu) = M_i(\omega,\nu) \cdot \prod_{i=1}^{n} T_i(\omega,\nu) \qquad (18)$$

where M_o (ω, ν) is the displayed modulation spectrum and M_i is the input modulation spectrum of the image, and T_i (ω, ν) is the modulation transfer factor of the system component i.

Beaton (1984) evaluated the MTFA, GSFP, ICS, EP, IC, and SN metrics using the image dependent form of the metric. He regressed the metrics on subjective performance data. The subjective task was the Imagery Interpretability Rating Scale (1978). Performance scores were first converted to z-scores to account for changes in scaling strategies for each of the different images. Correlations between subjective performance and each metric were low, accounting for only 48 - 58% of the variance. Beaton repeated the correlations using data that were collapsed across images. Results for this analysis are summarized in Table 7.

The MTFA gave the best predictive capability. The other metrics did not perform well. However, this should not rule out evaluation of these metrics with digitally compressed images. If the structure or nonuniformities in compressed images can be measured as noise, the SN metric may still provide predictive capabilities.

Table 7. Correlations between image dependent metrics evaluated by Beaton (1984) and subjective rating performance.

Image Quality Metric	R^2 Subjective Ranking
MTFA	0.85
GSFP	0.625
ICS	0.575
EP	0.25
IC	0.60
SN	0.375

Pixel-Based Metrics:

Techniques used by image processing research for evaluating image fidelity are metrics that seek to minimize the error variance between an original image and the coded image. A statistically-based method commonly used is mean square error (MSE). The MSE metric as well as other pixel-based metrics are image dependent and do not take into consideration the human visual system. However, they have been modified by researchers to take into consideration the observer as described below.

Mean Square Error (MSE):

The mean square error (MSE) metric, frequently used in digital image processing, measures the difference between an original and modified image and is defined as

$$MSE = \frac{\Delta\omega\Delta\nu\sum_{x=0}^{n-1}\sum_{y=0}^{m-1}[M_o(\omega,\nu)-M_m(\omega,\nu)]^2}{\Delta\omega\Delta\nu\sum_{x=0}^{n-1}\sum_{y=0}^{m-1}M_o(\omega,\nu)^2} \tag{19}$$

where M_o and M_m refer to modulation spectra of the original image and modified image. The visual system is sensitive to differences in intensities and to areas in images where there are abrupt changes in intensity (edges); that is, the intensity and the gradient are important. However, the MSE metric performs an averaging, weighing all errors equally independent of the intensity or gradient (Levine, 1985). Therefore, it is not surprising that the MSE metric does not correlate well with human performance.

Perceptual Mean Square Error (PMSE):

The PMSE attempts to take the visual system into account. The deviations in the MSE are weighted by the CSF. PMSE is mathematically defined as

$$PMSE = \frac{\Delta\omega\Delta\nu\sum_{x=0}^{n-1}\sum_{y=0}^{m-1}[C_l(\omega,\nu)[M_o(\omega,\nu)-M_m(\omega,\nu)]]^2}{\Delta\omega\Delta\nu\sum_{x=0}^{n-1}\sum_{y=0}^{m-1}C_l(\omega,\nu)^2 M_o(\omega,\nu)^2} \tag{20}$$

Beaton (1984) evaluated this metric and found a low correlation (0.15) between PMSE and subjective rankings.

Hall (1981) used a similar approach. Original images and degraded images were transformed using the Mannos and Sakrison vision model (1974). (This model is similar to the Hall and Hall model discussed in Section II). After transformation, the MSE between each degraded and original image was determined. The MSE was correlated with subjective performance resulting in a correlation of $R^2 = 0.92$.

Differences in results between these two studies may be due to experimental differences or differences in the visual models. Further investigation to compressed images are needed.

Contrast Energy Difference Metric (CED):

Farrell and Fitzhugh (1990) describe a contrast energy metric (CED) which is also based on differences. Original and modified images are weighted by the contrast sensitivity function then the squared differences of corresponding points in the original and modified image summed across each point in the image.

$$\sum_{i=1}^{N} (I_i - C_i)^2 \qquad (21)$$

where I_i and C_i refer to corresponding intensity points in the original and modified images respectively. This metric is also a sum of squared error, but it is not normalized with a mean.

In this experiment, the signal detection paradigm was used. Subjects were asked to discriminate between the original image and a modified image. In other words, identify which was the original and which was the modified image. Correct responses (hits) and incorrect responses (misses) are recorded. Two probability of occurrence distributions are determined. One distribution is considered a "noise" distribution (no signal present). The second distribution is "signal + noise." In this paradigm, d' is an indication of a person's sensitivity and is measured by the degree of separation between the two probability distributions expressed *in units of standard deviations*. For a review of this paradigm, readers are referred to Gescheider (1985).

Farrell and Fitzhugh (1990) compressed intensities of three capital letters (R, O and an ampersand) into 2, 4, 8, 16, 32, 64, 128, or 256 levels of grey. Subjects were asked to discriminate between original and modified images using the signal detection paradigm. Performance of the discrimination task (measured as d') was monotonically related to the log CED. Data were fitted using the Weibull psychometric function. The authors provide no explanation for using a log of the function, or for fitting data to a Weibull function.

The CED metric does not perform the averaging that the MSE metric does. However, this metric may be dependent on using discrimination tasks. Many tasks require extraction of information, not determination of differences between images.

Farrell, Trontell, Rosenberg, and Wiseman (1991) tested the CED metric using complex imagery, such as the photograph of LENA. They reported similar results, however, they also found that the metric did not perform equally well across different images. When two distinct images with the same CED value are presented, subjects do not perform equally well in terms of discrimination between original and compressed images. This result indicates that the metric will not perform well with changes in image content.

Pixel-based metrics have not been as successful as MTF-based measures of image quality at predicting human performance and less attention has been paid to these metrics by researchers. Snyder (1985) pointed out that these metrics are not supported by empirical vision research. Furthermore, they have not been evaluated for compressed images. New metrics need to be developed and systematically evaluated.

SECTION IV - EMPIRICAL MODELLING

The image quality metrics described above are based on a theoretical approach which details information about the images or imaging system and include quantitative information about the visual system through a visual model. An alternative approach is to develop a pool of possible image quality predictors and determine empirically which predictors define image quality. This approach has been taken by Snyder and Maddox (1978), Kuperman (1985) and Decker (1989).

For example, Decker (1989) investigated the effects of spatial luminance nonuniformities on perception. The nonuniformities were described in terms of spatial frequency, modulation, gradient shape, and dimension. The descriptions of the nonuniformities were regressed against subjective impressions of the nonuniformities. (Data were collapsed across all subjects). R^2 values of 0.84 were found.

Kuperman (1985) used a vision spatial frequency channel model approach and regression to develop a metric. Six aircraft images were filtered using seven Gaussian filters with different center frequencies and bandwidths of 1.5 octaves. Subjects were asked to provide interpretability ratings and confidence ratings for each of the six aircraft images filtered by each of the gaussian filters (42 images). Regression analysis was performed to predict interpretability ratings based on center frequency of the filter. An R^2 value of 0.562 was reported.

The technique of multiple predictors describing the image quality metric has not received enough attention. In addition to using the regression equation as the image quality metric, regression analysis techniques can be used for variable screening to determine what variables are important to image quality metric for specific task measures. Researchers have primarily used linear regression techniques because of their ease of use. Nonlinear regression techniques may be applicable. The primary drawback to this technique is that the models are often task and situation specific; however, this approach should not be ruled out.

SECTION V - HUMAN PERFORMANCE MEASURES

The purpose of image quality metrics is to provide a quantitative method for predicting human performance or human perceptions of images. The use of subjective impressions of quality or objective performance measures depends on the purpose for which the image is intended. For example, medical images require information extraction in which case a metric that correlates to objective performance is needed. For a television picture, subjective impressions of the image are enough for determining customer satisfaction. In most cases, metrics have been used primarily to predict subjective impressions. If subjective impressions correlate well with performance then it is feasible to use the subjective data. For example, the NATO scale for photointerpretations was found to correlate well with information extraction (Snyder, Shedivy, and Maddox, 1981). However, subjective measures do not always correlate well with performance, and subjective techniques are not very robust. A metric that is not dependent upon the performance measure would be ideal. However, researchers have not focused on investigating and determining good performance measures. In some cases, the ability of a metric to predict performance may be due to the fact that the performance measure does not have construct validity. That is, the measure is not tapping into the construct that it is intended to measure. Therefore, research investigating the various performance measures is necessary.

It should also be noted that the metrics developed to date model early stages of the visual process and do not include the higher level cognitive processing. Modelling these processes is difficult. To develop a reliable, predictable metric models of cognitive processing should also be investigated.

SECTION VI - RECOMMENDATIONS

There are two general categories of metrics, those that evaluate the entire MTF response of a system unrelated to the image content, and those which are image dependent. Image processing researchers use image dependent metrics which have not had the same success as image independent metrics. However, for development of a metric for digitally compressed images, the image dependent metrics may be appropriate because the effect of the display hardware is not the only factor contributing to the quality of the image. Published image quality metrics have not been systematically applied to digitally compressed imagery. Listed below are recommendations based on the literature review.

1. Apply published metrics to digitally compressed imagery. The MTFA, SN, and IC metric computed as image dependent metrics should be investigated and compared to MSE type metrics.
2. Evaluate nonlinear visual models (eg., Hall and Hall, 1971). If the compressed image is transformed through the visual model, then output from the model could be correlated with human performance data. Such an approach would be time consuming and more complicated than using the metrics described in this report. In addition, without adding a cognitive component to the models, results are unlikely to be much different than current metrics.
3. New metrics should be developed and evaluated that take into

consideration higher levels of cognitive processing.

4. Research should also be focused on investigating the performance measures that are to be correlated with metrics to determine if the correlations change as the performance measure is changed.

The lack of any new or unique research in the past 10 years indicates that there is a need for investigation into innovative approaches to the problem of image quality metrics.

REFERENCES

Air Standardization Coordinating Committee (ASCC). (1978). *Imagery Interpretability Rating Scale*, AIR STD 101/11A.

Beamon, W. S. & Snyder, H. L. (1975). *An experimental evaluation of the spot wobble method of suppressing raster structure visibility* (Tech. Report AMRL-TR-5-63). Wright-Patterson Air Force Base, OH: Air Force Aerospace Medical Research Laboratory.

Beaton, R. J. (1984). *A Human Performance based evaluation of quality metrics for hard-copy and soft-copy digital imagery systems*. Unpublished doctoral dissertation, Virginia Polytechnic Institute and State University, Blacksburg, VA.

Beaton, R. J. (1988). Linear systems metrics of image quality for flat-panel displays. In SPIE/SPSE *Electronic Imaging Devices & Systems Conference*.

Cohen, R. W. & Gorog, I. (1974). Visual Capacity: An image quality descriptor for display evaluation. In *Proceedings of the Society for Information Display, 15*, (pp.53-62). New York: Palisades Institute.

Coren, Porac & Ward (1984). *Sensation and Perception* (2nd ed.). San Diego: Harcourt Brace Jovanovich Publishers.

Decker, J. J. (1989) *Display spatial luminance nonuniformities: effects on operator performance and perception*. Unpublished doctoral dissertation, Virginia Polytechnic Institute and State University, Blacksburg, VA.

Decker, J. J., Pigion, R. D. & Snyder, H. L. (1987). *A literature review and experimental plan for research on the display of information on matrix-addressable displays*. (Technical Memorandum 4-87). Aberdeen Proving Ground, MD: U.S. Army Human Engineering Laboratory.

Farrell, J. E. & Fitzhugh, A. E. (1990). A discriminability metric for digital letterforms. In *Society for Information Display International Digest of Technical Papers*, XXI, (pp 25 -28). Playa Del Rey, CA: Society for Information Display.

Farrell, J., Trontell, H., Rosenberg, C., and Wiseman, J. (1991). Perceptual metrics for monochrome image compression. In *Society for Information Display International Digest of Technical Papers*, XXII, (pp. 631 - 634). Playa Del Rey, CA: Society for Information Display.

Gescheider, G.A., (1985). *Psychophysics, method, theory, and application* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Ginsburg, A. P., Cannon, M. W. & Nelson, M. A. (1980). Suprathreshold processing of complex visual stimuli: evidence for linearity in contrast perception. *Science, (208),* 619-620.

Hall, C. F. & Hall, E. L. (1977). A nonlinear model for the spatial characteristics of the human visual system. *IEEE Transactions on Systems, Man, and Cybernetics, smc-7(3)* 161-170.

Hall, C. F. (1981, August). Subjective evaluation of a perceptual quality metric. In *Proceedings of the Society for Photographic Instrumentation Engineers, Image Quality, 310* (pp. 200-204). Bellingham, WA: SPIE.

Hufnagel, R. E. (1965). *A search for a summary measure of image quality. Part II.* Paper presented at the Annual Meeting of the Optical Society of America, Philadelphia.

Kuperman, G. G. (1985). *Bandpass spatial filtering and information content* (Tech. Report AAMRL-TR-85-046). Wright-Patterson Air Force Base, OH: Harry G. Armstrong Aerospace Medical Research Laboratory.

Kuperman, G. G. & Wilson, D. L. (1991). *Objective and subjective assessment of image compression algorithms.* In Proceedings of the Society for Information Display, XXII, (pp. 627-630). Anaheim, CA: Society for Information Display.

Levine, M. D. (1985). *Vision in man and machine.* New York: McGraw-Hill, Inc.

Mannos, J. L. & Sakrison, D. J. (1974). The effects of a visual fidelity criterion on the encoding of images. *IEEE Transactions on Information Theory, 20*(4), 525-536.

Schade, O.H. (1953). Image gradation, graininess, and sharpness in television and motion-picture systems. Part III. The grain structure of television images. *Journal of the Society of Motion Picture and Television Engineers, 61*, 97-164.

Schindler, R. A. (1976). *Optical power spectrum analysis of display imagery. Phase I: Concept Validity* (Tech. Report AMRL-TR-76-96). Wright-Patterson Air Force Base, OH: Aerospace Medical Research Laboratory.

Snyder, H. L. (1980). *Human visual performance and flat panel display image quality* (Tech. Report HFL-80-1). Virginia Polytechnic Institute and State University, Blacksburg, VA.

Snyder, H.L. (1985). Image quality: measures and visual performance. In L.E. Tannas (Ed.), *Flat-panel displays and CRTs* (pp 70 - 90). New York: Van Nostrand.

Snyder, H. L. & Maddox, M. E. (1978). *Information transfer from computer-generated, dot-matrix displays* (Tech Report HFL-78-3). Virginia Polytechnic Institute and State University, Blacksburg, VA.

Snyder, H.L., Shedivy, D.I., & Maddox, M.E. (1981). *Quality metrics of digitally derived imagery and their relation to interpreter performance: III. Subjective scaling of hard-copy digital imagery* (Tech. Report HFL-81-2). Blacksburg, VA: Virginia Polytechnic Institute and State University.

Stockham, T. G. (1972). Image processing in the context of a visual model. *Proceedings of the IEEE, 60(7)*, 828-842.

Task, H. L. (1979). An evaluation and comparison of several measures of image quality for television displays. (Tech. Report AMRL-TR-79-7). Wright Patterson Air Force Base, OH: Air Force Aerospace Medical Research Laboratory.

Task, H.L. & Verona, R. W. (1976). A new measure of television display quality relatable to observer performance (Tech. Report AMRL-TR-76-73). Wright Patterson Air Force Base, OH: Air Force Aerospace Medical Research Laboratory.

Task, H. L. & Pinkus, A. R. (1987). Contrast sensitivity and target recognition performance: a lack of correlation. In *Society for Information Display International Digest of Technical Papers*, XVIII, (pp 1275 - 129). Playa Del Rey, CA: Society for Information Display.

van Meeteren, A. (1973). *Visual aspects of image intensification*. Soesterberg, The Netherlands: Institute for Perception TNO.

Wilson, H. R. & Bergen, J. R. (1979). A four mechanism model for threshold spatial vision. *Vision Research, 19*, 19-32.

☆U.S. Government Printing Office: 1992 — 648-069/40242